



А.Д.Журавлев, В.С.Нечунаев

## РАСПРЕДЕЛЕННЫЕ ВЫЧИСЛЕНИЯ НА ПЛАТФОРМЕ BOINC

(ФГБОУ ВПО «Самарский государственный аэрокосмический университет им. академика С.П. Королева (национальный исследовательский университет)»)

**Введение.** В современном мире перед человечеством стоит много сложных технических и научных задач. Один из способов их решения это компьютерное моделирование. Чтобы провести моделирование с высокой степенью достоверности, нужны огромные вычислительные мощности. Обычно для этих целей используют суперкомпьютеры либо GRID системы, из кластеров или серверов. Но есть другой способ – объединить ПК частных лиц в GRID сеть с помощью специального программного обеспечения. И об одной из таких программных платформ, а именно о BOINC (Berkeley Open Infrastructure for Network Computing) и пойдет речь в данной статье.

### **Добровольные распределенные вычисления**

Добровольные распределенные вычисления (ДРВ) – это распределенные вычисления на компьютерах, добровольно предоставляемых частными лицами, различным научным проектам. Участие в таких проектах не накладывает ни на пользователей ни на их компьютеры каких либо ограничений. Так как всё что для этого требуется это несколько простых условий. Первое - иметь доступ в Интернет для связи с сервером проекта. И второе - это простая установка клиентской программы, которая запрашивает задания, выполняет необходимые вычисления и отправляет полученные результаты на сервер проекта, где те проходят дополнительную проверку.

### **Платформа BOINC**

Платформа BOINC представляет собой стандартный набор из двух основных частей: серверная часть (BOINC Server) и клиентскую часть (BOINC Client) — оболочки, которую устанавливают у себя участники ДРВ. Серверную часть ПО, устанавливают и настраивают организаторы проектов, она включает в себя программные модули по созданию заданий, их рассылке и учету. Клиентская часть ПО находится на компьютере участника и представляет собой сам клиент, плюс удобный графический интерфейс (BOINC Manager). Он позволяет подключаться к различным проектам распределенных вычислений, контролировать получение заданий, ход расчетов, наличие или отсутствие ошибок и т.п. Первоначально платформа разрабатывалась для крупнейшего проекта добровольных вычислений — SETI@home, но впоследствии разработчики из Калифорнийского университета в Беркли сделали платформу доступной для сторонних проектов. На сегодняшний день BOINC является универсальной платформой для проектов в области математики, молекулярной биологии, медицины, астрофизики и климатологии. BOINC даёт исследователям возможность задействовать огромные вычислительные мощности персональных компьютеров со всего мира. По данным на октябрь 2012 года BOINC являет со-



бой распределённую сеть из более чем 540 000 активных компьютеров, со средней производительностью около 6600 терафлопс.

### **Серверная часть BOINC**

Серверная часть состоит из HTTP-сервера с веб-сайтом проекта, базы данных MySQL и набора специальных служб: генератор заданий, планировщик, службы для проверки и обработки результатов. Сервер устанавливается только на Linux, предпочтительно Debian.

Службы представляют собой набор программ на C++.

HTTP-сервер представляет собой набор PHP-скриптов и необходим организаторам проектов для общего управления проектом: регистрация участников, распределение заданий для обработки, получение результатов, управление базами данных проекта.

В базе данных хранятся пользователи, пароли, записи заданий, результатов, информация о хостах, программах проекта и прочее.

### **BOINC-клиент**

Для пользователей понятие BOINC чаще используется в контексте понятия BOINC-клиента — универсального клиента для работы с различными (BOINC-совместимыми) проектами распределённых вычислений. BOINC-клиент позволяет участвовать одновременно в нескольких проектах с помощью одной общей программы управления.

Для визуализации процесса управления BOINC-клиентом можно использовать либо поставляемую по умолчанию официальную программу-менеджер либо воспользоваться «неофициальной» программой для мониторинга и управления BOINC-клиентом. Следует отметить, что собственно BOINC-клиент в академическом понимании не имеет пользовательского интерфейса как такового, а представляет собой сервис, запускаемый при запуске системы, и управляется по протоколу TCP/IP. Однако, для конечного пользователя это не имеет значения, поскольку дистрибутив программы комплектуется программой-менеджером, которая сразу по умолчанию устанавливается вместе с BOINC-клиентом как единое целое и абсолютно прозрачна для пользователя. В этом случае в качестве адреса управляемого программой менеджером BOINC-клиента указывается адрес «localhost». Таким образом, с одной стороны, ничто не мешает пользователю использовать альтернативную программу-менеджер для управления BOINC-клиентом, а с другой стороны даёт возможность управлять несколькими BOINC-клиентами, находящимися на разных компьютерах из одной программы-менеджера. Также такая организация управления BOINC-клиентом подразумевает возможность использовать BOINC-клиент в «невидимом» режиме, когда запускается исключительно сервис, без пользовательского интерфейса вообще.

### **Наиболее известные сейчас ДРВ проекты**

Активных BOINC-проектов сейчас очень много, поэтому определить популярность довольно сложно. Но можно выделить группу из SETI@home, Einstein@home, World Community Grid, Rosetta@home, ClimatePrediction.net, MilkyWay@home и LHC@home.



SETI@home занимается обработкой данных, получаемых на радиоастрономической обсерватории в Аресибо — с целью поиска в «космическом шуме» необычных, искусственных сигналов, которые бы явно свидетельствовали о том, что мы не одиноки во Вселенной. Масштабность и необычность задачи привели к тому, что SETI@home стал на долгие годы самым популярным проектом.

Einstein@Home занимается поиском гравитационных волн и нейтронных звезд (пульсаров), обрабатывая данные, поступающие из трех обсерваторий различного типа: гравитационных интерферометров LIGO, радиоастрономических обсерваторий Аресибо и Паркс и космической гамма-обсерватории Fermi.

World Community Grid — по сути целый набор подпроектов, каждый из которых в рамках WCG может рассылать свои задания. Основные направления — фармакология и разработка технологий ресурсосбережения и экологии. Главным организатором этого проекта является гигант компьютерной индустрии — IBM.

Rosetta@home моделирует процессы свертывания белков (фолдинг) и поиск их трехмерных структур, соответствующих состояниям молекул с минимальной энергией по известной химической формуле белка, что необходимо знать для поиска новых лекарств.

ClimatePrediction.net моделирует климат как всей планеты Земля в целом, так и ее отдельных участков с учетом различного набора факторов, влияющих на него.

MilkyWay@home, «просеивая» данные Слоановского обзора неба, выделяет из общей массы те звезды, которые принадлежат к звездным потокам вокруг нашей Галактики и моделируют историю их развития.

LHC@home моделирует прохождение пучков частиц внутри Большого адронного коллайдера. Цель проекта — подбор оптимальных значений параметров более чем 1000 магнитов таким образом, чтобы пучок терял как можно меньше частиц во время работы ускорителя, а детекторы, соответственно, собирали как можно больше интересующих физиков данных.

### **Какие важные результаты уже удалось получить с помощью добровольных распределенных вычислений?**

Напомним, что 4 июля 2012 года учеными CERN было объявлено об обнаружении новой частицы, которая с высокой долей вероятности является бозоном Хиггса, поиски которого продолжаются уже почти полвека. Без проекта распределенных вычислений LHC@home открытие могло бы и не состояться или состояться значительно позже.

В Einstein@home открыто более трех десятков пульсаров, разработаны новые алгоритмы выделения полезных данных из зашумленного сигнала интерферометра LIGO для поиска гравитационных волн.

В MilkyWay@home удалось восстановить структуру приливного потока в созвездии Стрельца и реконструировать его историю.

В математических проектах типа PrimeGrid активно идет поиск простых чисел специального вида.



Российские проекты SAT@home и Gerasim@home уже выполнили свои первоначальные задачи по анализу алгоритма шифрования A5/1 и анализу качества разбиений граф-схем параллельных алгоритмов соответственно. А проекту SAT@home недавно удалось отыскать 2 пары новых диагональных ортогональных латинских квадратов порядка 10.

### **Кто организует проекты распределенных вычислений?**

Спектр организаторов очень широкий — от небольших коллективов и даже одиночек до крупных научных центров. Основная причина организации таких проектов — относительная дешевизна, поскольку большая часть вычислительных ресурсов проекту достается практически даром. Трудно представить стоимость аренды суперкомпьютера для проекта типа Einstein@home, работающего уже много лет и достигшего производительности более 800 терафлопс.

### **Что побуждает людей участвовать в такого рода проектах?**

Есть несколько причин. Одна из них это стать участником научного открытия. Так как участники, предоставившие ресурсы своих машин для вычислений обычно упоминаются при публикации результата. Например, в проекте Einstein@home публикуют имена (ники) тех участников, чьи компьютеры помогли обнаружить пульсары. Так же поступают организаторы SAT@home и многих других проектов. Следующая причина это желание помочь науке. Потом идут такие как: желание привлечь свой компьютер к решению полезных задач (ведь, по статистике, наши ПК большую часть времени не нагружают ни процессор, ни видеокарту серьезными вычислительными задачами); посоревноваться в статистике с другими участниками, командами или даже странами.

### **Литература**

1. Весь мир как суперкомпьютер / ТрВ № 110. – с. 7. / URL: <http://trv-science.ru/2012/08/14/ves-mir-kak-superkompyuter/>
2. BOINC URL: <http://ru.wikipedia.org/wiki/BOINC>
3. Платформа BOINC для организации распределенных вычислений. URL: <http://boinc.berkeley.edu/>
4. Российские распределенные вычисления на платформе BOINC URL: <http://www.boinc.ru/>